CHAPTER VIII

THE RELIABILITY COEFFICIENT FOR CROSS-COVARIANCE

by

J. Briscoe Stephens and Benjamin C. Hablutzel
NASA - Marshall Space Flight Center, Alabama

SUMMARY

The concept for the reliability coefficient of a covariance curve has been developed to improve the previous methods for identifying a peak through confidence levels which employ an arbitrarily chosen probability.

The technique described in this Chapter differs from the technique traditionally used for peak identification in that, rather than setting an arbitrary probability, the probability is now determined as a function of the statistical error and the covariance coefficient. For lack of a better term, this function has been defined as the reliability coefficient.

The physical significance and the reliability coefficient are discussed in terms of a stationary model. Results from the application of the reliability concept to actual atmospheric cross-beam data are discussed.

NOMENCLATURE

INDEPENDENT VARIABLES

Symbol	Definition
t	total observation time
T	total integration time
$\triangle T$	length of time for each piece -- the whole data record is subdivided into pieces for data processing.

Symbol	Definition
i	piece number
m	total number of pieces of data integrated over ($m = T/\Delta T$)
f	frequency
τ	time lag between the two signals
τ_m	maximum time lag

DEPENDENT VARIABLES

$x(t)$	signal time history for the sensor
$y(t-\tau)$	delay time history for the second sensor
$\overline{r_i(\tau)}$	covariance coefficient of the ith piece of data at time lag τ
$\overline{\overline{r(\tau)}}$	accumulative covariance average over the entire integration time
$\Delta r(\tau)$	statistical error of the covariance
$P(t_p,m,\tau)$	reliability for a time delay

OPERATORS

$( )$	statistic
$\overline{( )}_i = \frac{1}{\Delta T} \int_{(i-1)\Delta T}^{i\Delta T} ( )\, dt$	piecewise mean (Note: when average done over real time or lag, the symbol $< >$ is used.)
$<( )_k> = \frac{1}{k} \sum ( )_i$	accumulative mean where k is the kth piece of data (when averaged over pieces)
$\overline{\overline{( )}} = <\overline{( )}_{k=m}>$	total accumulative mean

Symbol	Definition
$\langle ( )_k \rangle^2 = \frac{1}{k-1} \sum_{i=1}^{k} (( )_i - \overline{( )_k})^2$	statistical error.

VIII-1. INTRODUCTION

The concept of a reliability coefficient for the evaluation of cross-covariance curves obtained from atmospheric cross-beam measurements is developed in this chapter. The reliability concept is a new way of viewing the traditional peak identification procedure, which uses an arbitrary probability level based on test results. The procedure described here requires that the maximum positive peak in the covariance curve (the mean covariance) be greater than the statistical error. When this stipulation is fulfilled, the time lag corresponding to this peak is defined as the transit time. This is the lapsed time used by the turbulence eddy in traversing the distance between the beams.

When an arbitrary probability level is used, validity of the correlation curves is obtained as a binary response and not as a relative measurement. Comparison of the statistical error at different time lags does not determine which time lag is more statistically significant. This fact means that the current approach does not afford a systematic means to compare different parts of the same correlation curve. Furthermore, the variability of the experimental parameters from test to test prohibits the use of marginal correlation results since the use of an arbitrary probability level requires an extra safety factor to account for all test conditions.

The reliability coefficient being proposed here affords a relative evaluation of the reliability of the correlation curve based on both the sample population and the statistical error. This relative evaluation results in the possible elimination of the limitations of the arbitrary probability levels when the reliability coefficient is used as the key criterion for evaluating the consistency of the correlation results. The same information can be obtained directly from the reliability probability used with arbitrary probability levels.

It is hoped that this approach can account for piecewise variations in both the transit time and the amplitude of the peak. It should be recognized that this is merely a proposed approach and not a solution to these problems.

This chapter assumes that the reader is familiar with both the correlation technique (Jayroe and Su [VIII-1]) and the atmospheric cross-beam program (Stephens, Sandborn, and Montgomery [VIII-2]).

VIII-2. THEORETICAL DEVELOPMENT

The statistical error of the covariance can be interpreted physically, as follows, rather than mathematically.

The accumulative covariance is defined as

$$\overline{\overline{r(\tau)}} = \frac{1}{N} \int_0^T [x(t) - \bar{\bar{x}}][y(t-\tau) - \bar{\bar{y}}]\, dt, \tag{1}$$

where x and y are the instantaneous fluctuations of the signal, $\bar{\bar{x}}$ and $\bar{\bar{y}}$ are the accumulative mean values, T is the total integration time, t is the real time measurement from the start of the record, and τ is the time delay.

The statistical error $\overline{\overline{\Delta r(\tau)}}$ for the covariance is

$$\overline{\overline{\Delta r(\tau)}} = \left\{ \frac{1}{(m-1)} \sum_{i=1}^{m} [\overline{r_i(\tau)} - \overline{\overline{r(\tau)}}_m]^2 \right\}^{1/2}, \tag{2}$$

where m is the number of pieces of data used, $\overline{r_i(\tau)}$ is the ith covariance occurring at the time delay τ, and $\overline{\overline{r(\tau)}}$ is the resulting accumulative covariance for the mth piece of data.

When using the statistical error for providing a confidence limit, it is required that

$$\overline{\overline{r(\tau)}} > \overline{\overline{\Delta r(\tau)}} \cdot t_{p=80\%} \tag{3}$$

for the covariance coefficient to be significant. This Student's "t" of 80%, $t_{p=80\%}$, implies a probability level of 80 percent. There are three

categories into one of which a covariance coefficient can fall by the scheme set forth by this last restriction. The covariance coefficient at a time lag can be less than, greater than, or equal to the statistical error.

When the statistical error is greater than the covariance coefficient, i.e.,

$$\overline{\overline{r(\tau_a)}} < \overline{\overline{\Delta r(\tau_a)}} \cdot t_{p=80}, \tag{4}$$

then we say that this is not an acceptable measurement. However, a measurement is statistically significant, i.e.,

$$\overline{\overline{r(\tau_b)}} > \overline{\overline{\Delta r(\tau_b)}} \cdot t_{p=80}, \tag{5}$$

if the average $\overline{\overline{(\ )}}$ exceeds its own standard error $\overline{\overline{\Delta(\ )}}$.

The third category is the one where the statistical error is equal to the covariance coefficient:

$$\overline{\overline{r(\tau_c)}} = \overline{\overline{\Delta r(\tau_c)}} \cdot t_{p=80}. \tag{6}$$

This relationship implies that there exists an 80 percent probability that the sample and the expected value of the covariance have the same sign.

Thus, when the covariance coefficient of a positive covariance sample equals the statistical error, then we have a relation for the probability of having a positive correlation coefficient. This relation is

$$\overline{\overline{r(\tau)}} = \overline{\overline{\Delta r(\tau)}} \cdot t_p, \tag{7}$$

where the probability level, p, is the unknown. However, this unknown is related to the known confidence factor

$$t_p = \frac{\overline{\overline{r(\tau)}}}{\overline{\overline{\Delta r(\tau)}}} \tag{8}$$

by the Student's "t" distribution.

The probability level required to satisfy the Student"s "t" relationship given in equation (8) is

$$P(m;\tau) = \frac{2 \cdot \Gamma(m/2)}{\sqrt{\pi(m-1)} \cdot \Gamma(\frac{m-1}{2})} \int_{x=0}^{x=t_p} [1 + \frac{x^2}{(m-1)}]^{-\frac{m}{2}} dx. \qquad (9)$$

The result for $P(t_p,m,\tau)$ will be defined as the "reliability coefficient." Although the reliability probability was developed for the cross-covariance, it is equally applicable to piecewise derivations of any other statistics. The Student's "t" distribution was used because the samples under consideration are normally small ($m < 30$). For larger piece numbers, the normal distribution might be used, i.e., the reliability coefficient becomes independent of the number of pieces used.

VIII-3. APPLICATION OF THE RELIABILITY TO THE STATIONARY COVARIANCE CURVE

The variation of the reliability coefficient with time lag can be used to afford a more meaningful peak identification technique because:

(1) The reliability curve provides a systematic procedure to establish the relative reliability of different portions of the same covariance curve.

(2) The reliability curve can be used for estimating the noise level tolerance.

(3) The reliability coefficient accurately accounts for the uncertainty associated with small sample size by employing the Student's "t" distribution.

Before discussing the properties of the reliability coefficient, we must first review additional characteristics of both the covariance and standard error.

The limitation in computer storage prohibits the integration of the covariance (equation (1)) over the entire record at one time. Therefore, the record is subdivided into m pieces at length ΔT such that

$$T = m \cdot \Delta T. \qquad (10)$$

For each piece, a piecewise covariance coefficient $\overline{r(\tau)_i}$, where i is the piece number, is obtained. Then the accumulative covariance coefficient $\overline{\overline{r(\tau)_m}}$ for the entire record is

$$\overline{\overline{r(\tau)_m}} = \overline{\langle r(\tau) \rangle_m} = \frac{1}{m} \sum_{i=1}^{m} \overline{r(\tau)_i}. \tag{11}$$

Thus, the statistical error $\overline{\Delta r(\tau)}$ defined in equation (1) is the effective measure of how the piecewise covariance $\overline{r(\tau)_i}$ varies about the accumulative covariance $\overline{\overline{r(\tau)_m}}$. The primary physical requirement of the correlation technique is that the phenomenon being measured is recurring. Thus, the consistency of certain piecewise covariance coefficients is the necessary property (implied by the requirements of the correlation technique) that enables us to retrieve temporal information. The reliability is nothing more than an estimate of this consistency.

To expand on the significance and application of this requirement of consistency, let us consider the following argument. This argument will be restricted to stationary data obtained from the convection of a frozen pattern at a constant speed such that the eddies traverse the transit distance between the two beams in a transit time, τ_s. The piecewise covariance coefficient at the time delay τ_s should be

$$0 < \overline{r(\tau_s)_i} < 1, \tag{12}$$

and will be referred to as "the covariance due to the signal." The remaining covariance coefficients, $\overline{r'(\tau_s)_i}$ (the prime indicates all time delays other than when $\tau = \tau_s$), can have values between

$$-1 < \overline{r'(\tau_s)_i} < 1, \tag{13}$$

and will be referred to as "the covariance due to the noise." (For ease of discussion, the signal component is defined as a spike in the covariance curve. This does not restrict the effect of this discussion.)

For random stationary data, the signal component of the accumulative covariance curve has the following property:

$$0 < \lim_{m \to \infty} \overline{\overline{r(\tau_s)}} = \text{constant} \leq 1, \tag{14}$$

while the noise component has the property that

$$\lim_{m \to \infty} \overline{\overline{r'(\tau_s)}} \simeq 0. \tag{15}$$

This property has the resultant implication that the peak in the covariance curve due to the signal remains at approximately the same time lag from piece to piece, but a peak in the piecewise covariance due to noise will be moving from piece to piece. Previous peak identification efforts thus assumed that the peak corresponding to the convection speed in the accumulative covariance curve can be differential from a peak of equal magnitude due to noise by using the statistical error of the covariance, since the statistical error of the signal peak is less than the statistical error due to the noise peak. A more sophisticated approach is required in practice because of finite integration times and because of the varying magnitudes of the peaks in the covariance curve.

The reliability concept affords an approach which should circumvent these problems. To understand the behavior of the reliability coefficient, let us consider the following (see eqs. (7) and (8)): A large reliability coefficient at a time delay means that the covariance coefficient has the same sign from piece to piece; whereas, a small reliability coefficient means that the covariance coefficient at that time delay is oscillating from piece to piece. These oscillations are the characteristics of noise and it is thus apparent that the signal should have a higher reliability than the noise, once a sufficient sample of the population has been obtained. Thus, the reliability curve affords a method that, along with the covariance curve, can permit the signal to be distinguished from the noise in the curve before the noise has been completely suppressed.

The above discussion can be extended to set a noise tolerance by using a knowledge of the beam geometry and the wind direction. With this knowledge, one can determine whether the transit time should be a positive or negative time lag. The signal half-plane will be defined as the half-plane of the covariance curve in which the transit time must exist. The other half-plane is the noise half-plane, since any peak existing in this half-plane will be assumed to be due to noise. The maximum reliability coefficient which exists in the noise half-plane is then a measure of the minimum acceptable confidence level, and will thus give the noise tolerance.

This approach obviously has some shortcomings, the major one being a periodic signal. For a periodic signal, the noise half-plane will have a reliability equal to the sum of the noise and the signal. Thus, in this case, a more complex formula will have to be used to obtain the noise level. An example of such a formula is to subtract the known signal reliability coefficient from the maximum reliability coefficient existing in the noise half-plane. The result is the noise reliability level. These problems should not exist with a broadband process, but it is always necessary to know the kind of process that is being measured. The above concepts are thus incomplete and should be taken as a basis for a new systematic approach to peak identification, not as a finished identification criterion.

To summarize the application of the reliability coefficient for setting the noise tolerance, let us consider the following two applications. First, for zero beam separation (Figure VIII.1), the procedure for the interpretation is as follows. The signal is what occurs around zero time delay. All other peaks then must be noise peaks. If the noise reliability coefficient is less (see peak (a)) than the signal reliability coefficient, then the covariance curve can be meaningfully evaluated; but if the noise reliability coefficient is greater (see peak (b)) than the signal reliability coefficient, then it does not seem to be meaningful to accept these results. (The case of a periodic signal has been omitted.)

The second application is to the measurement of a wind speed (Figure VIII.2). In this case, the signal range of time lags is set by the range of possible wind speeds during the period of observation. Once this is established, the same argument that was used in the case of a zero beam separation applies here. If the reliability coefficient of the signal is not greater than the reliability coefficient of the noise, then the results are questionable.

The reliability affords one final feature that can be useful for setting the integration times. When the reliability of the signal peak does not increase (or if it decreases) with additional integration time,

and if the probability of noise does not decrease, then additional integration time will afford no improvement in the covariance curve. The advantage of this over the accumulative error currently used is that it provides closer control. That is, the accumulative error monitors only the overall variation, while the reliability can be used to monitor specific peaks and to determine how the integration time affects these peaks.

VIII-4. APPLICATION OF THE PROBABILITY TO ACTUAL ATMOSPHERIC DATA

In section VIII-3 the discussion was limited to stationary data, because at best this approach leaves a certain amount to conjecture. When dealing with nonstationary data, the amount of conjecture increases; therefore, the power of these concepts will be empirically demonstrated.

The first case shown in Figure VIII.3 is a model example of the use of this probability concept. The peak in the covariance curve at -10 seconds compares favorably with the wind speed. The signal peak of the reliability coefficient is about 99 percent, while the maximum noise peak is only 73 percent. The fact that the reliability coefficient of the signal is greater than the reliability coefficient of the noise indicates that the signal peak is significant, a fact already obvious from the statistical error. This is the only wind measurement evaluated that had a 99 percent reliability coefficient for the signal. We have found by independent experiments of other statistical techniques that this is our most well-behaved run.

The second case of a wind measurement, shown in Figure VIII.4, demonstrates the use of the "noise" half-plane. Winds compare only at negative lags. The peak at -15 seconds corresponds to the wind speed, but the peaks at -25 seconds is also statistically significant, according to the statistical error. Our concept holds that the noise peak at 50 seconds has a reliability coefficient of 71.6 percent. This eliminates the peak in the signal plane at -25 seconds, since it has a reliability coefficient of only 70.2 percent. We are left with the signal peak of a 81.6 percent probability, which is what we desired.

The third example is a zero beam separation (Figure VIII.5). According to the statistical error, we should reject the results shown in Figure VIII.5, even though they are the desired ones, because they are not statistically significant according to the statistical error. When the reliability coefficient is applied, it is found that it peaks at zero time delay (out to ± 75 seconds). Thus, the reliability coefficient identifies the peak at zero as meaningful. However, this

identification also indicates that the probability for the identified peak to be positive is only 21 percent. Our confidence in the "accepted" result is thus still low.

VIII-5. CONCLUSIONS

At this time, only a preliminary conclusion is possible since a great deal more work is needed with the reliability coefficient before a conclusion can be drawn. At the time this concept was tested, the concepts discussed in section VIII-3 seemed to be valid for all of the cross-beam data available at that time. Furthermore, this concept seems to increase our ability to evaluate the covariance curve, and seems to work for both stationary and nonstationary data. Tests suggest that it also may be useful in evaluating derivatives and spectra. More work is planned with the latter.

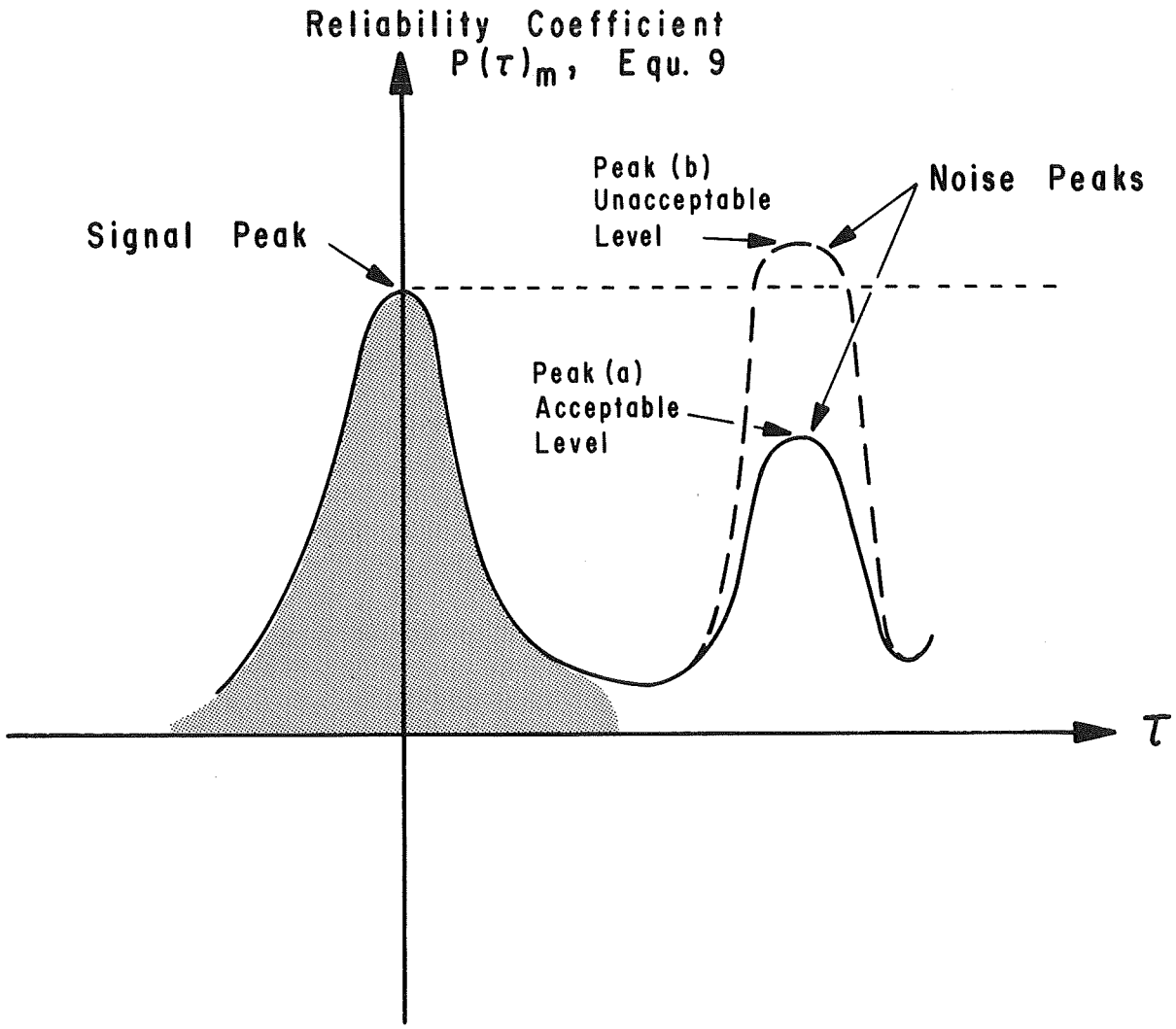

FIGURE VIII.1. MODEL RELIABILITY CURVE
FOR ZERO BEAM SEPARATION

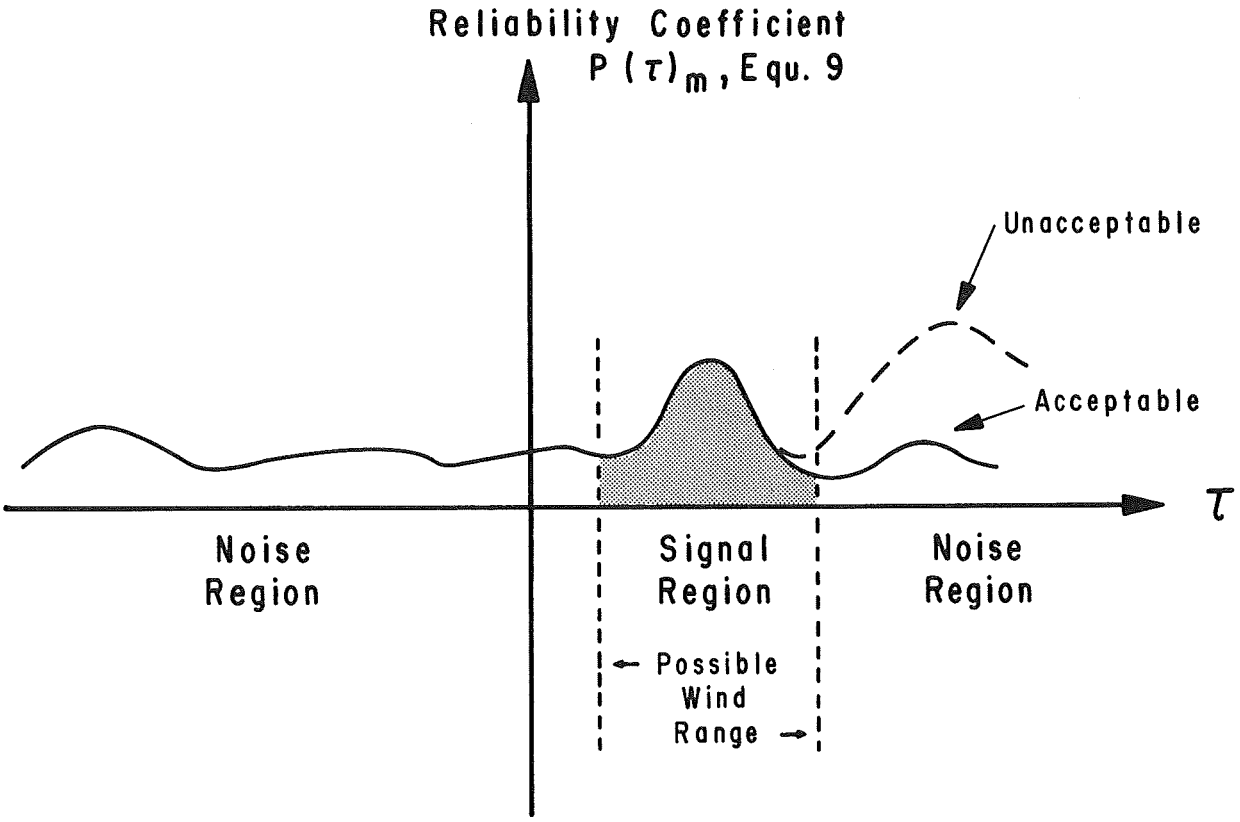

FIGURE VIII.2. MODEL RELIABILITY FOR A BEAM SEPARATION

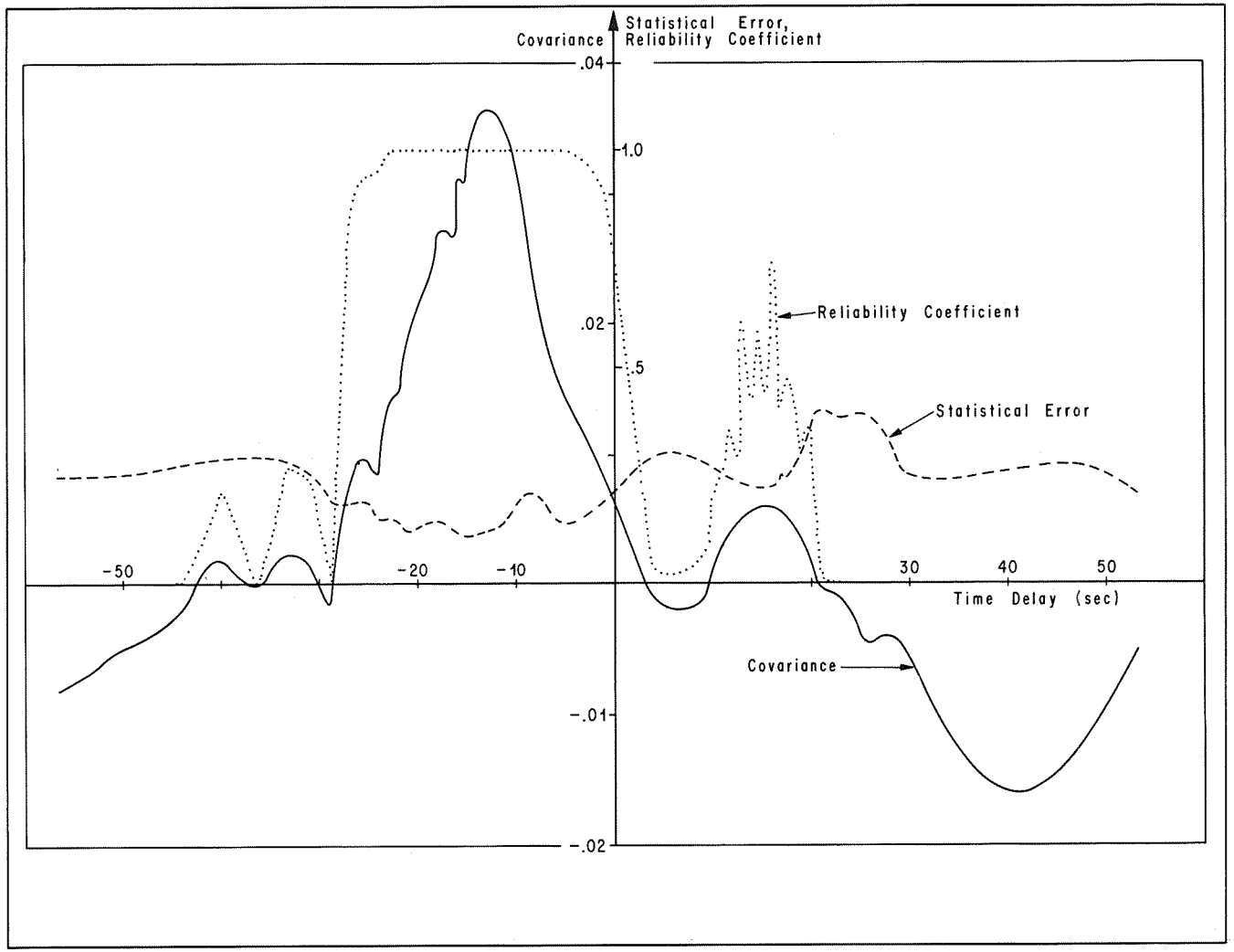

FIGURE VIII.3. ACTUAL BEAM SEPARATION

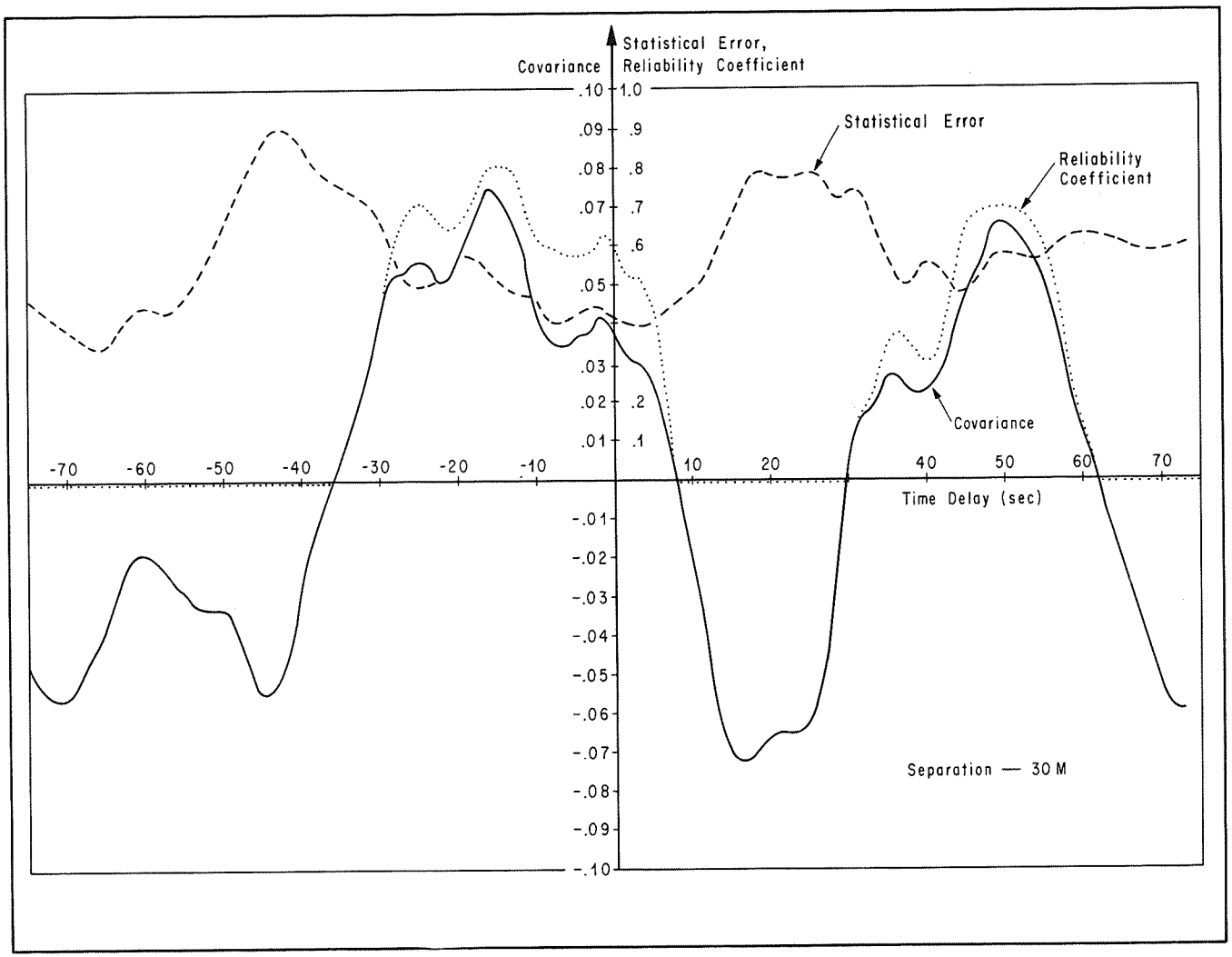

FIGURE VIII.4. ACTUAL BEAM SEPARATION

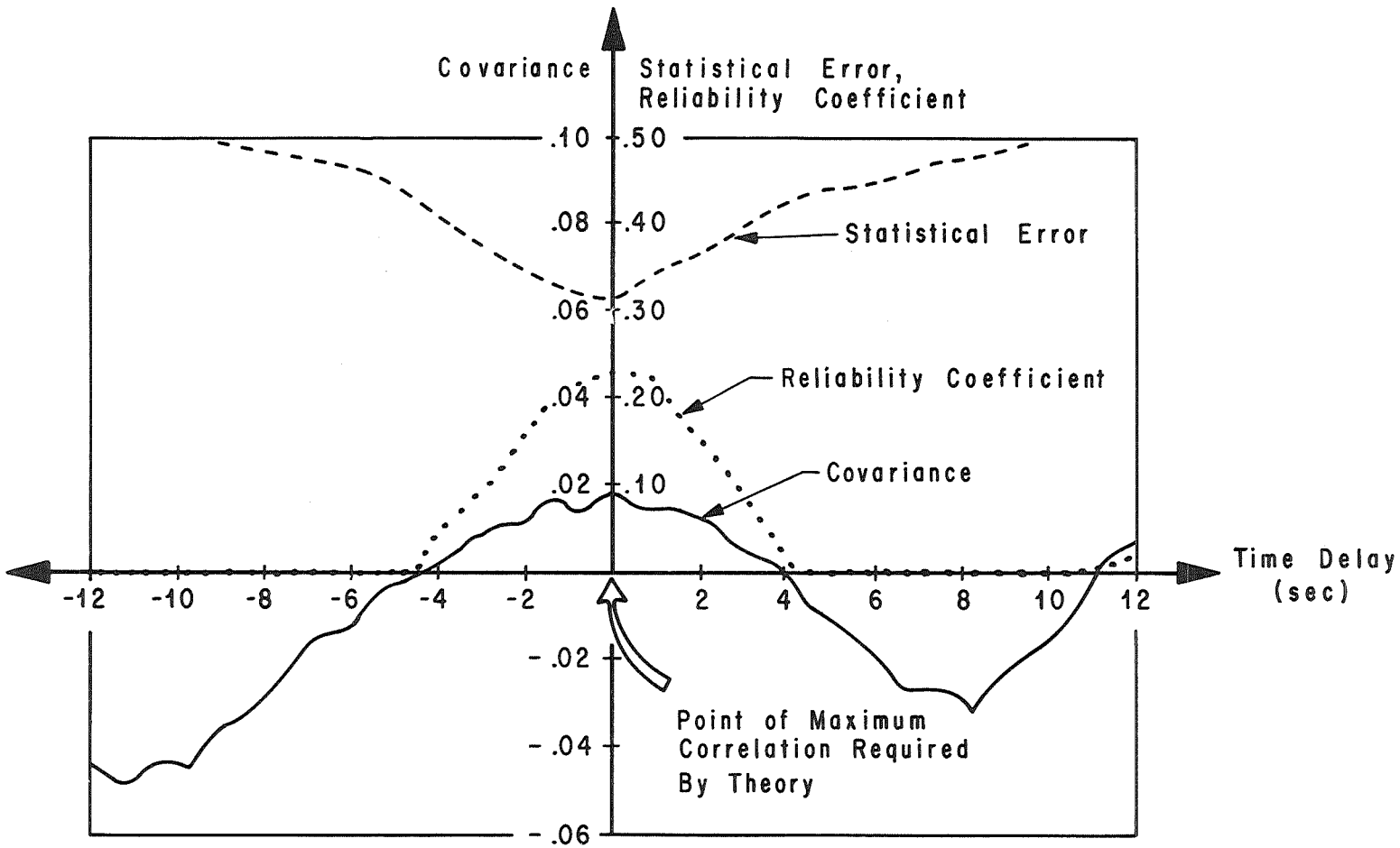

FIGURE VIII.5. ACTUAL ZERO BEAM SEPARATION

REFERENCES

VIII-1. Jayroe, R. R. and M. Y. Su, "Optimum Averaging Times of Meteorological Data with Time-Dependent Means," NASA TM X-53782, October 15, 1968, pp. 103-116.

VIII-2. "Remote Wind Detection with the Cross-Beam Method at Tower Heights," NASA TM X-53782.